We propose a sequential method for solving the multi-hypothesis testing problem in software reliability, and derive tests for the parameters involved under the Jelinski-Moranda, Shick-Wolverton, the geometric error detection rate, and the error-content proportional detection rate models. Sequential tests are obtained by applying a generalized Wald's sequential probability ratio test due to Bechhofer, Kiefer and Sobel (1968), and they guarantee an average probability of correct decision to be at least $1 - \beta$ (preassigned) when sampling terminates and a given terminal decision rule is applied. This new approach represents a natural application of classical results in sequential analysis to the (relatively new) area of software reliability testing.

1. Introduction

Software development and usage in all areas have become increasingly significant in recent years. The abundance and wide use of software has led to considerable research in software reliability, and several stochastic models have been proposed in the studies of software reliability theory. As a result, statistical inference problems in software reliability testing have become increasingly important. In most of the models proposed for studying software reliability, there are two basic parameters involved: (i) $N =$ the number of bugs (faults) in the system before testing starts, and (ii) $\lambda$ (or $\lambda(t)$) = the intensity parameter (or the intensity function) of the fault detection process. Statistical inference problems concerning those parameters have been studied under the Jelinski-Moranda model by several researchers. See, e.g., Joe and Reid (1985), Ross (1985), and Singpurwalla (1991).

In this paper we study the problem of multi-hypothesis testing of the parameters under the general framework of sequential analysis. In particular, we apply a fundamental theorem in the monograph of Bechhofer, Kiefer and Sobel (1968, Section 3.1) to develop procedures for sequential tests of $s$ ($s \geq 2$) hypotheses concerning the parameters $N, \lambda$ and parameters in other models. Specific stopping rules and terminal decision rules are obtained. Results concerning error probabilities and the expected sample sizes are given, and a unique property of the test procedures is that the average probability of a correct decision is bounded below by a predetermined constant $1 - \beta$. We believe that this is a first study on software reliability testing under the framework of sequential analysis.

In Section 2 we describe the Jelinski-Moranda and other related models in software reliability theory. Section 3 provides a brief description of the Bechhofer-Kiefer-Sobel Generalized Sequential Probability Ratio Test (denoted by BKS-GSPRT). We then apply the BKS-GSPRT in Section 4 to derive the sequential tests under the Jelinski-Moranda, Shick-Wolverton, and two discrete-time models. The properties of the tests are studied both analytically and empirically in Section 5.
2. Some Commonly-Used Stochastic Models in Software Reliability Testing

2.1. Continuous-Time Models

In error counting models with continuous time, a given software system (or program) is in operation continuously in time until a bug is detected. Once it is detected, it is then removed permanently from the system in a minimal amount of time without introducing new errors. The sequence of random variables observed during the debugging process is \( \{X_j\}_{j=1}^{N} \) where \( X_j \) is the waiting time to detect the \( j \)-th bug after the \((j-1)\)-th bug was discovered \( (X_0 \equiv 0) \). Under a given model, the joint distribution of \( X_N = (X_1, \ldots, X_N) \) can be determined.

2.1.1. The Jelinski-Moranda Model

Perhaps the most widely referenced model in software reliability is that proposed by Jelinski and Moranda (1972). In this model it is assumed that the random variables \( X_1, \ldots, X_N \) are independent and \( X_j \) has an exponential distribution with mean \( [(N-j+1)\lambda]^{-1} \) \((j = 1, \ldots, N)\) where \( \lambda \) is an intensity parameter. As a consequence, for every fixed \( k = 1, 2, \ldots, N \) the joint density function of \( X_k = (X_1, \ldots, X_k) \) is (for \( x_j \geq 0, j = 1, \ldots, k \))

\[
f_{X_k}^{(k)}(x_k) = \left[ \prod_{j=1}^{k} (N-j+1)\lambda \right] \exp \left[ -\sum_{j=1}^{k} (N-j+1)\lambda x_j \right],
\]

which, of course, depends on the parameters \( N \) and \( \lambda \).

The simplicity of the Jelinski-Moranda model makes it a particularly appealing one. But it assumes that all faults contribute equally to the system's failure. Boland, Proschan and Tong (1987) and others showed how fault diversity affects the performance of a software reliability test procedure when this assumption does not hold in certain applications.

2.1.2. The Shick-Wolverton Model

The Shick-Wolverton model is a simple modification of the Jelinski-Moranda model. Under this model the joint density function of \( X_k = (X_1, \ldots, X_k) \) is (for \( 1 \leq k \leq N \))

\[
g_{X_k}^{(k)}(x_k) = \left[ \prod_{j=1}^{k} (N-j+1)\lambda x_j \right] \exp \left[ -\frac{1}{2} \sum_{j=1}^{k} (N-j+1)\lambda x_j^2 \right].
\]

The models described in 2.1.1 and 2.1.2 can be treated as special cases of the general case in which the \( X_j \)'s are assumed to be independent random variables with Weibull distributions such that their parameters depend on \( N \) and \( \lambda \).

2.2. Discrete-Time Models

Another class of software reliability models involves discrete time. For \( t = 0, 1, 2, \ldots \) let \( Y_t \) denote the number of (new) bugs detected in the \( t \)-th test run \( (Y_0 = 0) \). Once those bugs are detected, they are removed permanently from the system without introducing new faults. For this class of models it is usually assumed that the sequence of accumulated numbers of bugs detected \( \{\sum_{t=1}^{k} Y_t : k = 1, 2, \ldots\} \), forms a nonhomogeneous Poisson process with mean function \( \psi_k = \sum_{t=1}^{k} \phi_t, k = 1, 2, \ldots \). Thus \( Y_1, Y_2, \ldots \) are assumed to be independent Poisson random variables with means \( \phi_1, \phi_2, \ldots \). Under this assumption the error detection rate per error per test run (which is the probability that a bug will be encountered during a test run) after the \( k \)-th test is given by \( \phi_{k+1}/(\sum_{t=k+1}^{\infty} \phi_t) = \phi_{k+1}/(\psi_\infty - \psi_k) \), where \( \psi_\infty \) indicates the expected number of bugs to be eventually detected. Since the number of bugs left in the system is a nonincreasing function of \( k \), in applications it seems reasonable to assume that \( \{\phi_t \}_{t=1}^{\infty} \) is a nonincreasing sequence.